

# 说明书

---

## 构建匹配抗原表位和抗体识别区及编码基因数据库的方法以及数据库、 存储介质和电子设备

5

### 技术领域

本发明属于生物技术领域，具体地说，涉及一种构建匹配抗原表位和抗体识别区及编码基因数据库的方法以及数据库、存储介质和电子设备。

### 10 背景技术

目前国内临床上主要通过化验血液肿瘤指标、影像学检查（B 超、CT、MRI、PET-CT 等）及近年来的风险基因检测、循环肿瘤细胞检测等方式检查肿瘤信息，但这些方法的敏感性和特异性均不高，发现有异常时往往已是中晚期。对于恶性肿瘤而言，早发现早治疗，才能有效的提高生存率。随着分子生物学和免疫学的不断深入研究，研究者已经认识到肿瘤的检测已上升到了抗原表位肽水平。而在现有技术中，针对一些肿瘤疫苗的开发，缺少相应的数据支撑，也缺少构建匹配抗原表位和抗体识别区及编码基因数据库的方法。因此，有必要提供构建匹配抗原表位和抗体识别区及编码基因数据库的方法。

20

### 发明内容

本发明的一个目的是提供一种构建匹配抗原表位和抗体识别区及编码基因数据库的方法的技术方案。

根据本发明的一个方面，本发明提供一种构建匹配抗原表位和抗体识别区及编码基因数据库的方法，包括以下步骤：

1) 对细胞样本进行第一次免疫图谱检查，明确细胞样本当前免疫能力，以此作为基线记录在案；

2) 使用肿瘤特异性集群靶标检测技术已有靶标库中每种特定抗原表位，

分别刺激细胞样本,使其产生特定适应性免疫能力,即包含特定 V(D)J 的 TCR;

3) 接受刺激一周后,对细胞样本进行第二次免疫图谱检查,将所得三维柱状图与基线对比,明显升高的即为该特定抗原表位的对应抗体识别区;

4) 进一步对含有此种 V(D)J 的细胞样本做基因测序,获得编码此种 V(D)J 的基因序列,以及构成此种 V(D)J 的氨基酸序列;

5) 对步骤 1) 至步骤 4) 中获得的数据整理记录;

6) 整合步骤 5) 中记录的氨基酸序列,得到特定抗原表位与特定抗原识别区及其编码基因的对应关系,构建形成匹配抗原表位和抗体识别区及编码基因的数据库。

10 可选地,还包括以下步骤:

对步骤 4) 中筛选得到的氨基酸序列删除重复,建立集合。

可选地,所述细胞样本包括淋巴细胞。

可选地,步骤 1) 中所述的免疫图谱检查的具体方法为:提取分离出的淋巴细胞的 RNA 或 DNA,构建文库后进行 NGS 测序。

15 可选地,步骤 6) 中对应关系具体为:

对特定抗原表位注释,注释信息包括能够产生相对应的适应性免疫能力的特定抗原识别区及其编码基因;

对特定抗原识别区注释,注释信息包括编码基因以及能够对哪些特定抗原表位产生相对应的适应性免疫能力。

20 可选地,步骤 6) 中对应关系还包括:将特定抗原表位与相对应的特定抗原识别区相互关联。

可选地,还包括以下步骤:

重复步骤 1) 至步骤 5),步骤 6) 对多次重复结果进行整合。

25 根据本发明的另一个方面,本发明还提供一种匹配抗原表位和抗体识别区及编码基因数据库,该数据库采用上述构建方法构建而成。

根据本发明的另一个方面,本发明还提供一种存储介质,所述存储介质内存储有上述的匹配抗原表位和抗体识别区及编码基因数据库。

根据本发明的另一个方面,本发明还提供一种电子设备,该电子设备包括上述的存储介质。

30 本发明的一个技术效果在于,通过本发明构建匹配抗原表位和抗体识别

区及编码基因数据库的方法，能够构建出抗原表位和抗体识别区及编码基因的对应关系，得到相应的数据库。得到的数据库能够为后期肿瘤疫苗开发和基因编辑提供数据支持。

通过~~以下参照附图对~~本发明的示例性实施例的详细描述，本发明的其它特征及其优点将会变得清楚。

## 附图说明

~~此处所说明的附图用来提供对本申请的进一步理解，构成本申请的一部分，本申请的示意性实施例及其说明用于解释本申请，并不构成对本申请的不当限定。在附图中：~~

~~图1是本申请一些实施例中免疫图谱检查分析得到的三维柱状图。~~

## 具体实施方式

以下将配合实施例来详细说明本发明的实施方式，藉此对本发明如何应用技术手段来解决技术问题并达成技术功效的实现过程能充分理解并据以实施。

根据本发明的一个方面，本发明提供一种构建匹配抗原表位和抗体识别区及编码基因数据库的方法，包括以下步骤：

1) 对细胞样本进行第一次免疫图谱检查，明确细胞样本当前免疫能力，以此作为基线记录在案。所述细胞样本通常需要包括淋巴细胞。具体的，该细胞样本的获取可以通过由献血者志愿献血的血站所采集、储存的血液获得，采血后 4 小时内 4℃ 冷藏运输至 GMP 实验室进行处理分离外周血淋巴细胞，保证样品及时有效的检测。提取分离出的淋巴细胞的 RNA 或 DNA，构建文库后进行 NGS 测序。最后使用独特的生物信息学分析算法对测序结果进行筛选比对分析，得到 VDJ 基因使用频率、CDR3 片段长度、克隆多样性分析等结果。免疫图谱分析是利用高通量测序技术来分析人的外周血或组织中 B 细胞的 BCR 序列或者 T 细胞的 TCR 序列，对疾病进行早期诊断筛查。人体内的 B 和 T 淋巴细胞是获得性免疫系统中两类重要的细胞。其

中, B 细胞在早期发育过程中表达细胞表面抗原受体 BCR, 在抗原和辅助因子刺激后表达成抗体(Antibody)分泌到细胞外; T 细胞表达细胞表面的抗原受体 TCR, 识别由 MHC 呈递在细胞表面的短肽抗原。BCR 和 TCR 中的 CDR3 部分都是由 V(D)J 基因重组产生的。V(D)J 基因重组发生在 B 或 T 淋巴细胞早期发育过程中。理论上讲, 每个 B 或者 T 细胞只表达一种重组过的 BCR 或者 TCR。BCR 和 TCR 的多样性是建立正常免疫功能的基础。但是, 在各种疾病中, BCR 和 TCR 的多样性可能被破坏。通过高通量基因测序的方法来分析 BCR 或者 TCR 基因的免疫图谱, 可以分析 B 或者 T 细胞的功能。分析结果可以以柱状图的形式体现, 现有技术中免疫图谱检查分析得到的三维柱状图参考图-1, 柱状图的位置及高度表明使用该 V $\beta$  和 J $\beta$  基因组合的 TCR 在本次检测中出现的频度。健康人群中, 每个健康个体可以使用多种 V $\beta$  和 J $\beta$  基因组合, V $\beta$  和 J $\beta$  基因的使用可以存在一定的个体偏好性。但是, 在亚健康或疾病人群中, V $\beta$  和 J $\beta$  基因组合使用可能表现为明显的单调性。这种情况如果长期持续存在, 提示免疫系统功能低下。

2) 使用肿瘤特异性集群靶标检测技术已有靶标库中每种特定抗原表位, 分别刺激细胞样本, 使其产生特定适应性免疫能力, 即包含特定 V(D)J 的 TCR。通过特定的刺激使其产生特定适应性免疫能力, 从而找到包含特定 V(D)J 的 TCR。

3) 接受刺激一周后, 对细胞样本进行第二次免疫图谱检查, 将所得三维柱状图与基线对比, 明显升高的即为该特定抗原表位的对应抗体识别区。若未产生特定适应性免疫能力, 则放弃该细胞样本, 重新选择新的细胞样本重复步骤 1) 至步骤 3)。

4) 进一步对含有此种 V(D)J 的细胞样本做基因测序, 获得编码此种 V(D)J 的基因序列, 以及构成此种 V(D)J 的氨基酸序列, 基因编码的测序为本领域的常规技术手段, 在此不再赘述。

5) 对步骤 1) 至步骤 4) 中获得的数据整理记录, 得到基础数据。

6) 整合步骤 5) 中记录的氨基酸序列, 得到特定抗原表位与特定抗原识别区及其编码基因的对应关系, 构建形成匹配抗原表位和抗体识别区及编码基因的数据库。

上述特定抗原表位与特定抗原识别区及其编码基因的对应关系具体包

括：对特定抗原表位注释，注释信息包括能够产生相对应的适应性免疫能力的特定抗原识别区及其编码基因；对特定抗原识别区注释，注释信息包括编码基因以及能够对哪些特定抗原表位产生相对应的适应性免疫能力。进一步的，对数据进行关联，将特定抗原表位与相对应的特定抗原识别区相互关联。

通过本发明构建匹配抗原表位和抗体识别区及编码基因数据库的方法，能够构建出抗原表位和抗体识别区及编码基因的对应关系，得到相应的数据库。得到的数据库能够为后期肿瘤疫苗开发和基因编辑提供数据支持。

在一些实施例中，对步骤 4) 中筛选得到的氨基酸序列删除重复，建立集合。某一氨基酸序列可能对应多种特定抗原表位，通过集合关联的形式使得数据更为集中，便于后期的关联检索以及分析。

在一些实施例中，重复步 1) 至步骤 5)，步骤 6) 对多次重复结果进行整合，对数据进行校验，避免特殊情况干扰导致得到的数据偏差，从而提高数据库的精准性。

树突状细胞集群靶标蛋白检测是一项精准的蛋白组学肿瘤筛查技术，通过血清检测完成针对 26 种肿瘤的 360 个特异性靶标的筛查，检测出阳性靶标的种类和数量，判断肿瘤的发生、发展、愈后与转归，实时监测指导临床治疗。该技术既是肿瘤早期发现的有效手段，又是肿瘤临床疗效评估的有效依据，同时也是指导个体化治疗、实现精准诊疗的基础。本发明使用该检测技术致力于找出特定抗原表位，与特定抗原识别区 V(D)J 及其编码基因的对应关系。从而为定向针对性的提高机体免疫能力，探索出理论依据和技术支持。远期，可以为肿瘤疫苗的开发，找到一条新的技术途径；还可以通过基因编辑技术改善人体免疫力。

根据本发明的另一个方面，本发明还提供一种匹配抗原表位和抗体识别区及编码基因数据库，该数据库采用上述构建方法构建而成。

根据本发明的另一个方面，本发明还提供一种存储介质，所述存储介质内存储有上述的匹配抗原表位和抗体识别区及编码基因数据库。具体的，该存储介质可以为硬盘等存储介质。

根据本发明的另一个方面，本发明还提供一种电子设备，该电子设备包括上述的存储介质。具体的，该电子设备可以是计算机、诊疗设备等具有上

述存储介质的电子设备。

还需要说明的是，术语“包括”、“包含”或者其任何其他变体意在涵盖非排除性的包含，从而使得包括一系列要素的商品或者系统不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种商品或者系统所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的商品或者系统中还存在另外的相同要素。

上述说明示出并描述了发明的若干优选实施例，但如前所述，应当理解发明并非局限于本文所披露的形式，不应看作是对其他实施例的排除，而可用于各种其他组合、修改和环境，并能够在本文所述发明构想范围内，通过上述教导或相关领域的技术或知识进行改动。而本领域人员所进行的改动和变化不脱离发明的精神和范围，则都应在发明所附权利要求的保护范围内。